APPENDIX B(1)

CRITIQUE OF HEW REPORT ON CIGARETTE SMOKING AND HEALTH CHARACTERISTICS

It is often desirable to ascertain some characteristic of the nation's population as a whole. For example the Census Bureau may want to know how many men and women there are in the country or what the average height and weight of each American is. Such information might be obtained by visiting or calling each individual in the country. However, the costs of such an approach would obviously be prohibitive.

The modern science of statistics has provided an alternative means of ascertaining this same information. Statistical methods, properly employed, permit conclusions to be drawn about the entire population from data collected from only a portion of the people. Thus when facts are repeatedly obtained from random samples of the population and subjected to statistical analysis it is possible to state with a known degree of certainty the extent to which these facts exist in the entire population. For example, on the basis of answers derived from repeated random samples we could state that it is more probable than not that the percentage of females in the United States is somewhere between 52.5% and 53.5%.

The enormous value of the statistical method is readily apparent and its uses have become widespread. However, this method poses a veritable quicksand for the unwary. Statistical methods improperly employed distort and misrepresent the actual facts thereby misleading the reader who lacks the training to perceive the subtleties inherent in the method.

Moreover, the possible sources of distortion are manifold. For example, unreliable opinions and not facts may constitute the data. The sample may not be sufficiently random, representative of the population as a whole or large enough. Or the data may not support the conclusions drawn.

These and many other sources of distortion may be introduced either inadvertently or intentionally. However, the greater the respect accorded to the results and the importance attached to them, the more concerned the statistician must be to prevent the reader from being misled. Where it is an agency of the United States Government which publishes results intending to affect the actions of the population as the whole, any distortion or bias (as it is referred to by statisticians) must be carefully avoided.

It is for this reason that the statistical deficiencies in the May 1967 report on cigarette smoking and

health characteristics by the United States Department of Health, Education, and Welfare are totally deplorable.

A. A Brief Summary of the Method Employed in the HEW Study.

The HEW report purported to compare the health characteristics of persons who are or had been cigarette smokers with persons who had never smoked. The basic methodology employed consisted of household interviews.

A sample of 42,000 households containing about 134,000 persons was selected. Each of these households was visited by an interviewer. If there were members of the household 19 years of age and over at home, the interviewer first asked those persons about their health particularly whether they had had any one of a number of chronic medical conditions such as heart conditions, chronic bronchitis, arthritis, hearing impairments and peptic ulcers in the last twelve months. The interviewer also asked questions related to the number of disability days suffered by the person being interviewed.

Second, the interviewer asked a series of questions related to smoking characteristics. Specifically each person was asked whether he had ever smoked 100 cigarettes in his life, how many cigarettes a day he had "usually" smoked when he was smoking the most, how many cigarettes a day he now smoked "on the average," how many cigarettes a

a day he was smoking twelve months ago "on the average," and how long it has been since he "smoked cigarettes fairly regularly." On the basis of the responses given persons were classified as former smokers, present smokers and never smokers. [1] Smokers were further classified by the number of cigarettes smoked per day--less than 11, 11 to 20, 21 to 40, and 41 or more.

The persons interviewed was also asked for this same health and smoking information for all members of the household who were not at home at the time of the interview, provided that they were related to the person being interviewed. These so-called proxy responses constituted the source of the data for 60.0% of the males who had ever smoked and 17.2% of the females who had ever smoked.

In addition, 5% of the households selected for sampling did not give any response either because no one was at home or because they refused to respond. The individua in these non responding households were imputed the characteristics of the persons in the same sampling segment who did respond.

1/ Of the males in the "never smoked" category 59.6% had smok cigars or pipe tobacco. Eighteen percent were smoking at the time of the survey. Of the male former cigarettes smokers 25.7% were cigar and/or pipe smokers at the time of the survey. This makes the male "never smoked" group quite different from the female "never smoked" group.

On the basis of the data collected in this manner conclusions were drawn as to the relationship between cigarette smoking and health characteristics for the entire population.

B. The Plainly Inadequate Sample.

It is a principle of elementary statistics that regardless of the size or randomness of a single sample, the results obtained from that sample may not necessarily be close to the results for the population as a whole. The most that can be said is that if the results of many such samples were averaged, their average will approach the average of the population.

On the other hand if even a single sample is inadequate, lacking sufficient size or randomness, the results of such a sample are worthless for any purpose--except perhaps to illustrate that poor results are obtained with inadequate samples.

In the case of the HEW report the sample was improperly designed and not large enough to support the conclusions drawn. This improper sampling casts a statistical cloud over the validity of all of the conclusions in the report.

The data in the HEW report is treated as if the sample consisted of 134,000 persons, the number of persons for whom responses were received or imputed. In addition estimates of sampling error were apparently based upon a sample of 134,000 persons.

The difficulty is that the sample was not adequately designed to constitute a truly random sample of 134,000 people. One of the requirements of such a sample is that each of the items of the sample must be completely independent of each other item.[2/] In the case of the HEW study each person was not independent of each other person. Instead whole families were interviewed. These are not likely to be independent in either smoking habits or health characteristics.

Similarly the independence of each sampling unit was further undercut by the fact that the non responding households, which comprised 5% of the sample, were imputed the characteristics of persons in other households who did respond. Also obtaining proxy responses further affected the independence of the individuals in the sample.

Moreover, the other major defect in the sample employed in the HEW study is that the sample was simply too small to permit some of the conclusions to be validly drawn.

2/ Arkin and Colton, An Outline of Statistical Methods, (4th Edition) 114.

In general, the size sample which one employs is simply a function of how certain one wants to be about the results obtained. Thus utilizing a large sample permits results to be expressed with a much greater degree of certainty than a small sample. However, there is still some minimum point for sample size in all situations. In the event that the sample size falls below this minimum, any results obtained from the sample are worthless. Thus for example, the average income in Washington, D.C. could hardly be approximated by averaging the incomes of three individuals living in Washington even if those individuals were selected at random.

It is this defect of plainly inadequate sample size which is present in many of the categories of the HEW report. For example, in Table 7 conclusions are drawn as to the number of females in the whole United States who presently smoke more than two packages of cigarettes a day and have a chronic condition. The data given in the report does not state how many females were interviewed who presently smoke more than two packages of cigarettes a day. However, by employing those figures which were given, it appears that the number of females actually interviewed who presently smoke more than two packages a day is approximately

200.[3] Moreover approximately 35 of these 200 were not interviewed directly but were the result of proxy responses.[4] In addition some of these females may have been in the same family and therefore have common characteristics. This would further reduce the actual number of random responses.

To suggest that any conclusions can be drawn as to all females in the United States from actual interviews with less than 200 females is patently absurd. One needs no statistical training to realize the inherent deficiency in such a miniscule sample size, and the misleading nature of any generalizations based upon it. Common sense will do.

C. The Inherent Unreliability of the Data.

Regardless of the care with which a sample is designed or the astuteness with which manipulations and

---

3/ This figure of approximately 200 females was calculated by first taking the total number of respondents, 134,000, and dividing by 2 on the assumption that 50% of the respondents were females. From figure 1 on page 2 it was observed that 33% of these females, or 24,000, were present smokers. Then from figure 2 on page 3 it was observed that 0.9% of all females who presently smoke, or 216 females, smoke more than two packages a day. [The figure of 200 was independ arrived at by Mr. Ramm's statisticians by a different method of calculation. In the event that we employ this figure in any further work, we can check their method of calculation in order to ascertain whether it is more precise than the one described above.]

4/ This was based on the fact that 17.2% of the female smoker had proxy respondents.

calculations are performed, any statistical inference is only as good as the measurements or data obtained. For example, if three people are asked to estimate the distance to the moon with their naked eyes, little reliance could be attached to the average of the three estimates regardless of how closely the three estimates agreed.

In the case of the HEW report the data is simply unreliable with respect to the prevalence of certain medical conditions and with respect to smoking patterns--the two basic characteristics being measured.

The medical data is unreliable because the source of these medical facts was not the physician--the only individual trained to diagnose. Instead each individual who happened to be home when the interviewer appeared became his own diagnostician informing the interviewer of the chronic conditions he had or thought he had. Moreover, this data was rendered still less reliable because of the widespread use of proxy responses which permitted individuals who were at home at the time of the interview to diagnose the chronic conditions of other members of the household.

The reporting biases which resulted from this approach of estimating chronic conditions by household

interviews--largely by proxy in many cases--are apparent. At best the person being interviewed could only describe conditions which an attending physician had diagnosed and passed on to the family. More likely, many individuals being interviewed went beyond this reported medical knowledge and played the role of a physician. The HEW report itself recognizes this widespread bias pointing out in the case of "Bronchitis and/or Emphysema" that

> "bronchitis as reported in the household interview is not necessarily the same as a physician's diagnosis of bronchitis. It is quite possible that some of the self-diagnosed cases may in fact be 'smokers' cough.'" (P. 12)

Other individuals may have welcomed the opportunity to tell someone what ailed them, and many wives may have expressed a sincere though medically unfounded concern for their husbands' health. These and many other biases make the medical data obtained highly unreliable. In apparent recognition of these biases the author of the HEW report found it necessary to state that,

> "the prevalence estimates produced from interview data will, for some conditions, <u>differ markedly</u> from estimates based on medical reports or clinical examinations." (P.5)

The author's candor in recognizing the unreliability of the data is commendable. Totally deplorable, however, is the

subsequent use of this data to draw conclusions about health which are offered to the public in an unqualified form.

Similarly the data on smoking characteristics is equally unreliable. Facts were not solicited--only opinions as to the amount and frequency of smoking. The interviewer did not ask, for example, the number of packages of cigarettes purchased yesterday or last week. Instead individuals interviewed were asked to give averages and amounts usually smoked--figures which involve subjective opinions.

Notwithstanding the reporting bias in such opinions, the report treats the data as facts when indeed they were not. An individual's ability to accurately describe his own present and past smoking patterns was undoubtedly affected by many factors such as his memory and whatever social significance he may have attached to the answer.

Even more important is the enormous reporting bias introduced into the data on smoking characteristics where proxy responses were solicited. This fact cannot be too strongly emphasized since data on 60.0% of the males and 17.2% of the females who had ever smoked were obtained from other persons. Individuals simply cannot report whether

another member of the family presently smokes 20 or 21 cigarettes a day "on the average." Certainly individuals do not know the smoking habits which someone else had during that person's entire life. Moreover, the proxy response given is plainly colored by whatever bias the individual asked has toward smoking.

Because of these and other factors the author of the HEW report concludes that

> "Self-and proxy-response differences also result in some reporting biases." (P. 5).

The report also notes that in fact the self-responses on the whole differed from the proxy responses.

In view of the recognized unreliability of the data, the authors of the report were not justified in drawing conclusions as to the number of cigarettes smoked a day within the ranges set forth in the report. Moreover the grouping of smokers by number of cigarettes smoked is particularly misleading since there is absolutely no measure of how long an individual smoked at a given rate or what his total exposure to cigarette smoking has been. Thus an individual who smoked two packs a day for one year would be placed in the same category as another individual who smoked two packs a day for twenty years. Notwithstanding this fact the report clearly implies

that total cigarette smoking exposure is being related to health.

It does not take a statistician to recognize that unreliable data cannot yield reliable conclusions regardless of the complexity of subsequent manipulations or the number of computers used to analyze the data. In the case of the HEW report a little thought about the methods employed for collecting the data and the application of simple common sense convinces the reader that both the medical prevalence data and the smoking characteristic data are so unreliable as to render completely unreliable any conclusions based upon this data. To publish such conclusions for unqualified public consumption is plainly misleading.

D. The Lack of Statistical Support for the Conclusions.

On the basis of the data collected, the HEW report draws numerous conclusions as to the relationship between cigarette smoking and the prevalence of various medical conditions for the population as a whole. However, a closer examination of the data given and the manipulations performed upon this data make it clear that these conclusions are completely unfounded for the population as a whole. In fact the conclusions are not

even warranted for the limited group of 134,000 persons comprising the sample.

As indicated above, it is a principle of elementary statistics that regardless of the size or randomness of a single sample the results obtained from a single sample may not necessarily be close to the results for the population as a whole. If a different sample of 134,000 persons were subjected to the same interviewing process, the results might differ markedly from those reported by HEW. The most that can be said is that if the results of many such samples were averaged, their average would approach the true value for the population.

Notwithstanding this indisputable statistical principle the HEW report not only draws conclusions as to the population as a whole based upon a single sample, but this report even tabulates its results only on a total population basis. The report does indicate in the "Technical Notes on Methods" (p. 56) that different households were interviewed each week and that each week's interviews constituted a separate sample at least for characteristics of high incidence. However, the report does not make it clear that there was an averaging of means for such high incidence characteristics and no probable errors or ranges of uncertainty

are given for any such specific values. In fact absolutely no group comparison statistical analysis is reported. In the absence of such information the report is not justified in drawing conclusions for the population as a whole even if results obtained were averaged.

Moreover as to the less common characteristics and smaller categories, the report itself implies that there was but a single sample. As indicated above, the results obtained from such a sample clearly do not permit the drawing of conclusions for the population as a whole.

This deficiency is no way remedied by the probable error calculations which can be made from the graphs on pp. 60-61. The fact that one value may have a probable error of X does not mean that another value has this same probable error. The second value may have a probable error many times larger than X. In short there is no justification in the report for reaching any conclusions as to the population as a whole.

Moreover on the basis of the data given the conclusions drawn by the HEW report are not even warranted for the 134,000 persons in the actual sample. One of the most elementary principles of statistics is that no

estimate of a value should ever be given unless it is accompanied by information which makes it possible to judge the reliability or precision of that estimate. 5/

The reason underlying this principle is perfectly clear. The result obtained from even repeated sampling is only an estimation of the true value. Unless it is known how close to the true value this estimation is likely to be, the estimate has no significance whatsoever. It is like informing a student that he scored 55% on an examination without telling him what the average score was. On some tests 55% may be the highest grade in the class; on others it may be a failing grade. The number standing alone has no meaning.

The authors of the HEW report failed to observe this basic principle. The report does not give any standard errors or other estimates of reliability for any specific values. It is therefore impossible to tell from the data and the charts which differences between values are statistically meaningful.

Thus for example Table 5 of the HEW report indicates that among all males who ever smoked the prevalence of chronic bronchitis and/or emphysema is

---

5/ Freund, Modern Elementary Statistics (3d ed.) 221-222.

2.3 per 100 males. The prevalence for persons who never smoked is given as 1.0 per 100. However, the value 2.3 per 100 might only mean that we are 95% certain that the true value lies somewhere between .3 per 100 and 4.3 per 100. Similarly the value of 1.0 per 100 might only mean that the true value lies somewhere between .1 per 100 and 1.9 per 100. In other words the prevalence rate for never smokers might actually be greater than the prevalence rate for smokers. The data given does not even establish that this is unlikely.

The HEW report does provide charts from which standard errors, reflecting sampling errors, may be calculated for groups of values. The author's intention is to provide a single standard error which could be applicable to a wide variety of health statistics. The author admits that such average standard errors do not indicate the precise standard error for any specific values. By necessity such an average standard error implies that certain values have much larger standard errors than the average and hence are much more unreliable than other values. The difficulty, however, is that it is simply impossible to tell from the data given how large some of the standard errors are and to which figures they are applicable. In short it is impossible to tell which

figures may be reliable and which are unreliable.

Quite apart from the failure to give specific estimates of the standard error for specific figures in the report, it is even likely that a careful evaluation of the reliability of the data was not carried out. The author suggests that "a number of approximations" were made. (P. 58.) However, we do not know how the graphs from which these average standard errors are to be calculated were prepared. These may have been based upon data obtained in other sampling work and might therefore be completely inapplicable to the present study.

Moreover, these average or general standard errors do not under any circumstances give a true measure of the accuracy of the data. At most these standard errors give some estimate of the sampling errors involved. They do not reflect in any way the many other obvious sources of error or bias in the data, some of which were discussed in Section C of this critique.

Nowhere in the report is any attempt made to estimate the magnitude of possible errors. Therefore anyone who may have assumed that values which have a small standard error are accurate has been totally misled.

In addition even if the standard error graphs are employed in the recommended manner to calculate the standard errors, it becomes clear that in many cases the standard errors are sufficiently large that the differences between the health characteristics of smokers and those who do not smoke are not statistically meaningful. To the extent this is true the failure to give any standard error information to the reader, without requiring him to resort to detailed calculations, is plainly misleading.

Thus, for example, Table 5 and Table 7 indicate that all chronic conditions are more prevalent among male smokers than among males in the total population. These tables indicate that the prevalence for male smokers is 122.4 per 100 males. For total males it is only 119.3 per 100 males. In point of fact there is no statistically meaningful difference between these two prevalence figures.

As the report points out any value given is only an estimate of the true value. The most that can be said is that there is a 68% chance that the true value lies within the interval between the estimated value plus one standard error and the estimated value minus one standard error.

When the standard errors are calculated for the prevalence of chronic conditions among all males and among male smokers, it turns out that the standard error for the male smokers is 1.9 and 1.6 for all males. Therefore the most that can be said is that there is a 68% chance that the true value for male smokers lies somewhere between 120.5 per 100 and 124.3 per 100. On the other hand for total males there is a 68% chance that the true value lies somewhere between 117.7 per 100 and 120.9 per 100. Therefore since these two intervals overlap, there is no statistically meaningful difference between the two values.

When this same type of statistical analysis is applied to the prevalence rates for any specific chronic conditions such as heart condition or bronchitis the results are even more startling. Since the prevalence rate figures given in the tables are based upon so few individuals in a given category reporting the condition, even the standard errors calculated from the graphs given on pages 60-61 are enormous. As a result any difference between figures which seem reasonably close such as 2.3 per 100 and 2.7 per 100 is almost certainly statistically meaningless. Even the difference in Table 5 between the prevalence of males who

ever smoked with chronic bronchitis and/or emphysema, 2.3 per 100, and the prevalence of males with chronic bronchitis and/or emphysema who never smoked, 1.0 per 100, is statistically meaningless.[6/]

Another illustration of the statistical uncertainty of the results is the fact that in the mere determination of the number of males smokers who ever smoked more than two packages of cigarettes a day the relative standard error is 3.2%. This means that there is only a 68% certainty that the true number of such smokers is somewhere between 2,667,000 and 2,843,000. However, it should be remembered that this 3.2% error merely includes an approximation of the sampling error in the determination of the number of male smokers smoking two or more packages a day. It does not include any of the biases or non-sampling errors involved in determining smoking characteristics. Nor does it include any biases or errors in obtaining the medical data. When this 3.2% figure is viewed in terms of these other sources of error, it seems perfectly clear that the conclusions being drawn and publicized for smokers who smoke more than two packages a day have absolutely no support in the data.

6/ Within the 68% certainty given the 2.3 is somewhere between 1.4 and 3.2; the 1.0 is somewhere between .1 and 1.9.

In some instances the authors of the report actually manipulate the data in such a way as to give an unfair picture of the results obtained. Conclusions involving percentages are unduly dramatized by the careful choice of a base. For example, the report states that among women who smoked 21-40 cigarettes a day at their heaviest rate, the number of those with three or more chronic conditions is 51% higher than that for "never smokers." (P. 8.)

This 51% figure was arrived at by comparing the 20.2% of female never smokers with three or more chronic conditions with the 30.6% of 21-40 cigarette a day female smokers having three or more chronic conditions. (See Table A.) Using the "never smoker" percentage as base, the percent difference of the two percentages is 51% figure.

Now, if statement of the problem is reversed, and expressed in terms of females who do not have three or more chronic ailments, we find 79.8% of "never smokers" and 69.4% of the 21-40 cigarette smoking group in this category. Then the percent difference of the percentages is 13% or 15%, depending upon choice of base, instead of 51% for the same data!

Since the difference between 20.2% and 30.6% may not be statistically meaningful, emphasizing this difference by manipulation of figures and percentages is plainly unwarranted. 7/

Notwithstanding all of these statistical deficiencies the authors of the report concluded that both male and female smokers reported higher rates of chronic conditions than persons who had never smoked, and this conclusion was widely publicized. However, when the actual collected data is examined, the data indicates that a smaller percentage of female smokers had one or more chronic conditions than females who did not smoke. (Table 3.) Similarly, all chronic conditions were less prevalent among females who smoked than females who had never smoked (Tables 6 and 7). In the case of males it has already been demonstrated that the data showed no meaningful statistical difference between the prevalence of chronic conditions among male smokers and among total males in the population. These facts, plainly indicated by the data actually obtained, would seem to totally destroy the validity of the conclusions reached in the report.

However, the authors of the report employed a further statistical manipulation to circumvent this result.

7/ For an interesting discussion of how percentages may be employed to mislead see Huff, How to Lie with Statistics, pp. 81-85.

The data was "age adjusted." Furthermore we are told that because of the small sample size it was necessary to do this age adjusting by the indirect method. After age adjusting in this fashion the authors concluded that the estimated percentage of persons with chronic conditions is greater for smokers than for the total population, both for males and females.

This procedure of age adjusting constitutes an admission by the authors that the sample was not representative of the population as a whole contrary to the position taken elsewhere in the paper. If indeed the sample was not representative with respect to age then the sample must also have been unrepresentative with respect to other factors relevant to health and smoking characteristics such as economic status and the proportion of persons living in rural as opposed to urban areas. Adjusting merely for age and not for these other biases is certainly of questionable statistical validity. Moreover, it is perfectly clear that using the indirect method of age adjusting introduces a further source of error. Therefore when this further source of error is considered in conjunction with the sampling errors, measurement errors and other biases and errors affecting the accuracy of the

results, there is considerable doubt as to whether any of the conclusions of the report are statistically sound or valid.

One final comment should be added on the question of causation. The author of the HEW report points out and quite properly that under no circumstances can data of the type given in that report establish any causal relationship between cigarette smoking and the various health characteristics discussed (P. 5).

CONCLUSION

In view of the serious statistical deficiencies in the HEW report and the study upon which it was based, the conclusions reached in that report were wholly unjustified. The wide publicity given to these conclusions without accompanying qualifications served only to mislead the population. Such conduct is improper when engaged in by private organizations. Coming from an agency of the United States Government it is intolerable.